Hyper-resolution monitoring of urban flooding with social media and crowdsourcing data

Wang, Ruo Qian; Mao, Huina; Wang, Yuan; Rae, Chris; Shaw, Wesley

Published in:
Computers and Geosciences

DOI:
10.1016/j.cageo.2017.11.008

Publication date:
2018

Hyper-resolution Monitoring of Urban Flooding with Social Media and Crowdsourcing Data

Ruo-Qian Wang[1,2], Huina Mao[3], Yuan Wang[4], Chris Rae[5], Wesley Shaw[5]

1 Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720
2 School of Science and Engineering, University of Dundee, Dundee, UK DD2 1BW
3 Oak Ridge National Laboratory, Oak Ridge, TN 37831
4 Department of Civil and Environmental Engineering, Tufts University, Medford, MA 02155
5 Blue Urchin LLC, 419 11th Ave E, Seattle, WA 98102

Abstract

Hyper-resolution datasets for urban flooding are rare. This problem prevents detailed flooding risk analysis, urban flooding control, and the validation of hyper-resolution numerical models. We employed social media and crowdsourcing data to address this issue. Natural Language Processing and Computer Vision techniques are applied to the data collected from Twitter and MyCoast (a crowdsourcing app). We found these big data based flood monitoring approaches can complement the existing means of flood data collection. The extracted information is validated against precipitation data and road closure reports to examine the data quality. The two data collection approaches are compared and the two data mining methods are discussed. A series of suggestions is given to improve the data collection strategy.

1. Introduction

Urban flooding is a global problem that costs lives and money. In 2010 alone, 178 million people suffered from floods. The total economic losses in 1998 and 2010 both exceeded $40 billion (Jha et al., 2012). Urban floods can be caused by a variety of reasons, including natural hazards of river overflow, coastal storm surge, sea-level rise, flash floods, groundwater seepage, sewer overflow, lack of permeability, and lack of city management. As urbanization proceeds and climate change intensifies, urban planners and city managers are facing the challenge of preparing for and mitigating flood damage. They need tools to monitor and predict the event for emergency response and development planning.

Monitoring and predicting urban floods needs high-resolution data with good coverage. High-resolution data can capture the variation of flood flows among streets or parcels, so that the heterogeneity of flood flows caused by heterogenous urban landscape can be captured. In this study, we define data that can reflect the variation on the parcel and street scale as "hyper-resolution" data. In addition to resolution, it is important to have a good coverage of flood data to obtain complete information.

The traditional method of obtaining flood related data lack both resolution and coverage. Remote sensing is a commonly used data source. Aerial photography, for instance, is being conducted by many research teams, engineering companies, emergency response services as well as governmental departments, and has demonstrated its value (Marcus and Fonstad, 2008). However, the systematic application of aerial photography is limited by vegetation canopies and cloud cover during floods (Hess et al. 1995, 2003). The limitation also exists in radar and satellite imagery that is based on optical sensors (Wilson et al., 2007). The most realistic and feasible approach is considered to be microwave remote sensing, which penetrates cloud cover. However, because of the corner reflection principle (Rees, 2001) along with coarse ground resolution, this technology is currently unable to extract flood data from urban areas. Another commonly used data source is distributed sensors. In the U.S., sensors have been installed in coastal areas by National Oceanic and Atmospheric Administration (NOAA) and in rivers and canals by United States Geological Survey (USGS), but almost no sensors are distributed on streets that are dedicated for urban flood monitoring purpose.

There are very few existing urban flooding datasets that can be used for detailed model validation and urban planning. A dataset about a flood event on January 10, 2005, in the City of Newport Beach, California, was built by a collection of 85 digital photographs and eyewitness reports from city employees who were dispatched to manage and photo-document the flood. The data collection involved interviews and email communications (Gallien et al., 2011). Another example in UK showed that contemporary newspaper reports can enrich the evidence of the witness reports and photos taken on the day following the storm (Smith et al, 2012). In addition, insurance reports can provide some additional information of flooding events such as damage evaluation. However, recording of such information is not a priority for civil defense agencies during a major coastal disaster, and data collected using this method could be expensive to access, incomplete in coverage, and substantially delayed.

Since urban area usually has a dense population, a big data approach, which relies on volunteered data reports from citizens, could be a potential solution to provide primary or complementary urban flooding data. The history of using social media data to study natural hazards can be traced back to Muralidharan et al. (2011), who compared the difference between non-profit organizations and media in the use of Twitter and Facebook during the Haiti earthquake in 2010. Since then, a series of studies based on social media data have emerged with a focus on floods. Sun et al. (2013) mentioned their effort to use Flickr images to support remote sensing based flood maps. Jongman et al. (2015) explored the potential to use Twitter data for early detection of flood events in Philippines and Pakistan. Fohringer et al. (2015) was the first to use flooding water depth information, which was manually extracted from the photos posted in Twitter. Eilander et al. (2016) studied the floods of Jarkata in Indonesia with Twitter data. They introduced a probability map to quantify the data uncertainty and found the general accuracy of location is around 69% but rises to over 90% if the specific location was mentioned in the content of the posts.

These studies clearly showed the value of data mining in flooding research, but their application is still limited due to the poor resolution of data collected. For example, the resolution of Jongman et al. (2015) is at city level. To achieve sub-city resolution, researchers heavily relied on manual reading as used in Sun et al. (2013) and Fohringer et al. (2015). The best resolution based on automatic algorithms was probably obtained by Eilander et al. (2016). Based on a group of independent tweets reporting flood depth, they constructed a probability map on the scale of urban community with an average size of a few square kilometers. To obtain the location information, the authors used manually defined location names and gazetteers to match location mentions in tweets. To support disaster monitoring, relief responses, model validation, and decision making, we still need higher data quality and accuracy to fully understand the details of flooding events.

The present paper is aimed at demonstrating a new approach to collect and process data for urban flooding research using Natural Language Processing and Computer Vision (CV) techniques. These techniques are shown promising to extract hyper-resolution data with a wide coverage to support urban flooding issues.

2. Data Source

2.1 Twitter

Twitter data was streamed using Twitter API during the period from September 29 to October 28 2015 and with the filtering keywords of “flood”, “inundation”, “dam”, “dike”, and “levee”. Retweeted posts labeled by Twitter were excluded. There are 7,602 tweets obtained.

2.2 Crowd-sourcing photos

We employed the crowd-sourcing platform MyCoast to collect photos about urban flooding. MyCoast is a system that, since 2013, has been used by a number of US environmental agencies to collect “citizen science” data about various coastal hazards or incidents. The app is built using Ionic, a user interface framework based on the Cordova cross-platform app framework. The web and data storage component is based upon WordPress and the app communicates with the server via a custom REST API. Data is stored in a MySQL back end. The app interface has been designed to be intuitive to those unfamiliar with phone apps, as the citizen scientist users often skew towards the youth or elderly end of the age spectrum.

There are two ways to contribute to the database, i.e. via the web and a mobile app. The user interface of the app is shown in Figure 2. The system currently contains over 5,000 flood photographs, and most of the photos were collected through the mobile app.

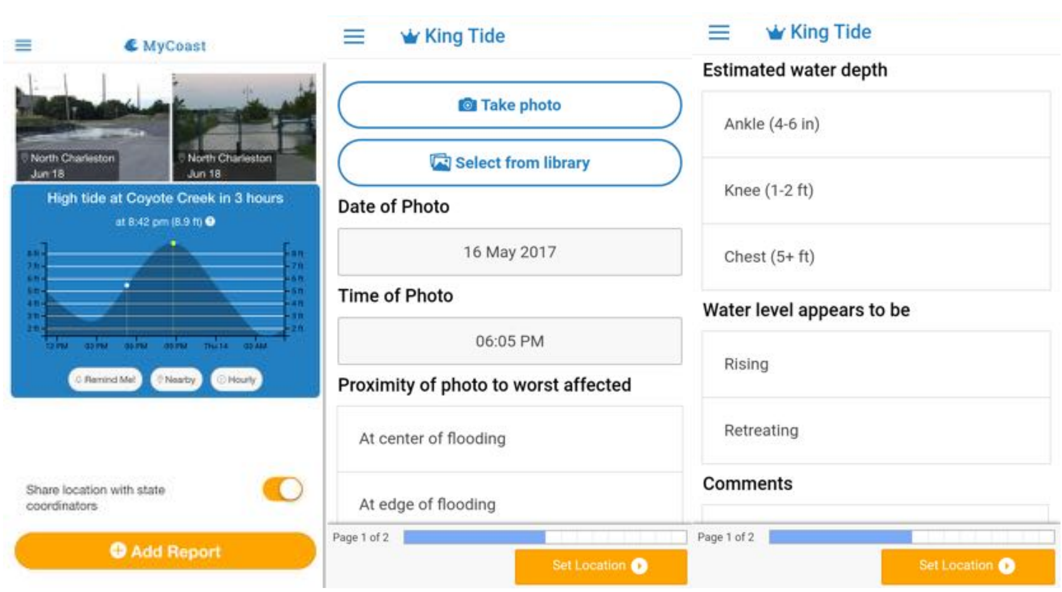


Figure 1 User interface design of MyCoast.

The app uses the phone's sensors to establish location and date/time information. Users are prompted to take a photograph and then may optionally add written comments (Figure 2). Users are also shown a chart with tide timing so that they can try to optimize the timing of photographs with peak tides.

2.3 Authorized Data

In the scale of the United States, a reliable data related to floods is precipitation statistics. The monthly precipitation data of October 2015 was downloaded from Advanced Hydrologic Prediction Service (AHPS), a database developed by National Weather Service (NWS) (https://water.weather.gov/precip/download.php). The precipitation departure of this month was the difference of the observation from the monthly normal precipitation obtained by Parameter-elevation Regressions on Independent Slopes Model (PRISM) using the record from 1981 to 2010. More details can be found in the NWS website.

Road closure information was collected using local media reports of September 29, 2015. We identified two major data sources, including the reports from WCBD online news at http://counton2.com/2015/09/29/coastal-flooding-closes-streets-downtown-second-day-in-a-row/ and the Twitter posts by Ashley Rae Yost, a reporter for WCBD television station in Charleston, SC. Integrating the data points, we can list the flooded roads on the day in Table 1.

Table 1 Collection of road closure data from media reports.

Name of the road	Starting point	End Point
Lockwood Drive	Gadsden St	N/A
Wentworth Street	Lockwood Blvd.	Gadsden St.
I-26	N/A	N/A
Lockwood Blvd	Beaufain Street	N/A
Broad St.	Barre St.	N/A
Barre St.	Beaufain	Montague
Lockwood St	N/A	N/A
Calhoun St.	Fourth St.	Halsey Blvd.
Market St.	Church St.	East Bay Drive
Long Point Road	Marsh Area	Marsh Area

3. Methods

3.1 Natural Language Processing methods

Location information is crucial for mapping floods. However, geotagged tweets are only about 1% of all the twitter data (Middleton, 2014), which poses a challenge for mapping. Even though the tweets are geotagged, the geotag coordinates may not be the same as the location of mentioned floods. For accurate location mapping, we aim to extract location mentions within the tweets' text. In addition, we extract the quantifier information, e.g. the depth of floods, which can help us understand the damage level of flooding. Table 2 shows several examples of tweets containing flooding location and depth information.

Table 2 Sample flooding tweets that are processed by NLP (location names in bold and depth numbers are underlined).

Tweet_ID	Posted_time	Tweet
648973656161394688	2015-09-29 14:32	*Roanoke River over 12 feet at **Walnut St bridge** in **Roanoke** now expected to top 13. Flood stage is 10. #swvawx",*
649935518038405120	2015-10-02 06:14	*Flood currently on **3 NE Wrightsboro** . in New Haven , NC . 1-2 ft . of water on Tandem CT . #ncwx #flood #flooded #flooding #rain #HurricaneJoaquin*
650019989353852928	2015-10-02 11:50	*Chicod creek at flood stage in **PItt County** , Likely to 11 ft by tonight . If it hits 12, it's running over the road.*
650393844115243008	2015-10-03 12:35	*Helped nearby drivers by reporting a flood on **PR-1, Ponce** on @username - Drive Social .*
650403125698891776	2015-10-03 13:12	*The **Market Street** portion of **Water Street** in **Downtown Wilmington** is under 9 inches of water*

In order to extract location names from tweets, we use the named entity recognition methods. Named Entity Recognition (NER) is a fundamental task in NLP, which aims to classify words into different types of names such as person, organization, location. Stanford's NER (https://nlp.stanford.edu/software/CRF-NER.shtml) is one of the cutting-edge named entity recognition tool available that uses Conditional Random Field (CRF) based classifier (Finkel et al., 2005). CRF is a conditional sequence model: given observations, it aims to find the highest possible sequence of states. In NER, the observations refer to words, and states refer to named entity tags. The entity tag for each word in a sentence is not predicted independently, but by considering the tags of neighboring words.

The original Stanford NER tool is trained on formal texts such as news data, which is remarkably different from short, informal, and noisy twitter data. So, applying the pre-trained NER model can generate poor performance in analyzing tweets. Lingad et al. (2013) have shown that NER tool retrained on Twitter data can significantly boost the performance on location detection from tweets. Therefore, we retrained the Stanford NER model using annotated tweets.

Data preprocessing: We pre-processed the data as follows. First, we remove non-english tweets (the python library "*guess_language"* is used to identify if a given tweet is English or not.) After filtering, 6,593 English tweets remain. Second, url links and mentioned usernames in tweets are replaced with unique words, "<URL>" and "@username", respectively. Third, each tweet is tokenized into as a list of words, which is fed into NER tool for location recognition.

NER model training and testing:

The training and testing data is provided by the ALTA 2014 Twitter Location Detection shared task. The original training and testing sets include 2,000 and 1,000 tweet ids, respectively. Since some tweets have been deleted or become invalid, we have obtained 1,851 in training and 930 tweets in testing. The training data is randomly split into 1,600 for training and 251 for validation. The annotated locations include place names and point-of-interests (POIs), in either main text strings or hashtags.

We have compared the performance of the original Stanford NER model and the re-trained model based on Twitter data. Results are evaluated in three metrics: precision, recall, and F-score, which are defined as:

Precision = TruePositive / (TruePositive + FalsePositive)

Recall = TruePositive / (TruePositive + FalseNegative)

F-Score = 2 × Precision × Recall / (Precision + Recall)

So, precision is the number of positive (or correct) results divided by the total number of all returned results (i.e. results that are predicted as relevant, which include both true positive and false positive samples), while recall is the number of positive (or correct) results divided by the total number of the actual relevant results. F-score combines precision and recall.

As it can be seen in Table 3, F-score is significantly improved after retraining. Even though original Stanford NER has a high precision rate, the recall is low (32.07%). In other words, a large number of location names are missed by this method. We have found that since the original NER model is trained based on the formal news data, it cannot capture many location information expressed in tweets, which are considered as informal language (e.g. abbreviations, misspellings, hashtags). Therefore, we use the re-trained Stanford NER model to detect locations from our flooding tweets.

Table 3 Results of location detection in ALTA 14 datasets.

Models	Precision	Recall	F-score
Stanford NER	94.51%	32.07%	47.88%
Stanford NER retrained	**86.68%**	**69.72%**	**77.28%**

Geocoding: From above, we have obtained a list of location names extracted from tweets. At this step, we need to geocode location names into geographical coordinates. Bing Maps Location API is used via Python's geocoding library (*geopy*), which accepts a location string as the input, and generates address name and latitude/longitude coordinates. For example, *"Battleship Rd"* identified from a tweet is converted to a standard address as (Battleship Rd, Camden, SC 29020, United States, (34.26021, -80.62693)). If more than one result are found by Bing Maps, the one with the highest probability is selected.

Flood depth information extraction: in order to capture the water depth mentions, we define regular expression patterns to capture numbers followed by "ft", "feet", "inch", "in" and their plural forms.

3.2 Computer Vision

The present study employs the technique of Convolutional Neural Networks (CNN) to automatically classify the crowdsourcing photos (Zeiler and Fergus, 2014). Originally inspired by animal visual perceptron, CNN is often used to detect and recognize objects. A good example of

CNN algorithm is Clarifai, which has won the competition of ImageNet Large Scale Visual Recognition Competition in 2013 (ILSVRC-2013). We applied this code through Clarifai API, which can be accessed as a remote web service. This service receives feedings of individual photos and feedbacks a list of tags that describe the objects present too busy to clean in the photo. For each tag, a probability is assigned to quantify its confidence of the recognition. Three examples of the recognition can be found in Figure 2 and the corresponding processed results are shown in Table 4.

Figure 2 Sample photos of the crowdsourcing data. Each photo is labeled by coordinates and time information as shown below the corresponding photo. The first photo shows a flooded street with trapped cars, bus, and trucks. The second photo has a water paddle at a port, but no spreading flood was observed. The third photo shows a river filled by water, but no flood can be recognized.

Table 4 Sample results of the computer vision applied to the sample photos shown in Figure 2.

Photo 1		Photo 2		Photo 3	
Tag	Probability (P)	Tag	Probability (P)	Tag	Probability (P)
flood	0.9964	no person	0.9678	flood	0.9961
rain	0.9963	wood	0.9656	water	0.9958
water	0.9940	furniture	0.9493	calamity	0.9904
tree	0.9834	chair	0.9381	storm	0.9816
river	0.9804	seat	0.9314	house	0.9796
reflection	0.9787	room	0.8956	building	0.9645
storm	0.9745	luxury	0.8942	river	0.9572
road	0.9696	industry	0.8807	architecture	0.9561
canal	0.9686	house	0.8705	no person	0.9530
no person	0.9652	water	0.8282	seashore	0.9426

Flood detection:

Flood is considered to happen if the flood tag was labeled. Applying this criterion to the photos in Figure 1, we found that flood was detected correctly in the first photo and no flood was correctly detected in the second, but a flood tag was incorrectly labeled to the third, which is a river in its normal condition. A manual check of 80 photos found that the accuracy is 65%, which is comparable to a text-image correlation study of social media (Chen et al., 2013) and an application study in earthquake damage estimation (Nguyen et al., 2017).

Visualization: The crowd-sourced data can be shown in Google map using the GPS coordinates contained in the labels of each photo. The road closure data were manually relocated, being visualized by a line connecting the starting and end points of the road closure section. Because the end points are missing in some road closure records, we only marked the starting points in these cases.

4. Results

4.1 Case study 1: the flooding map of the United States

The daily volume of flood related Tweets of the United States is shown in Figure 3. The daily volume increases quickly from October 1^{st} to October 4^{th}. This trend is consistent with the history of the Hurricane Joaquin, which reached the Category 4 major hurricane on October 1^{st} and gained its greatest strength on October 3^{rd}. Interestingly, the daily volume reached its top on October 4^{th}, a day later than the greatest strength of the hurricane. This phenomenon might be attributed to the reason that the flood event needed time to develop before flushing into the residential areas. Note that this one-day delay is also reported in the Philippine floods (Jongman et al., 2015). Immediately after the peak, we observe a downturn, which may be partially due to the incomplete data streamed from the Twitter API. Despite the incompleteness, this downturn was also observed by Jongman et al. (2015) in two of the Philippine floods, but missing in other reports, such as Eilander et al. (2016). We suspect that people were too busy cleaning up the disaster and the hurricane became of less value in news reports after diminishing.

The daily volume and the percentage of the hyper-resolution tweets that contained specific location are shown in Figure 3. The hyper-resolution tweets generally follow the trend of the total flooding related Tweets, and the average percentage is around 51%. This percentage is higher than expected and might indicate that most of the hyper-resolution tweets were posted by authorized agencies. Note that another hurricane developed on October 24, so the Tweeting volume had a second peak.

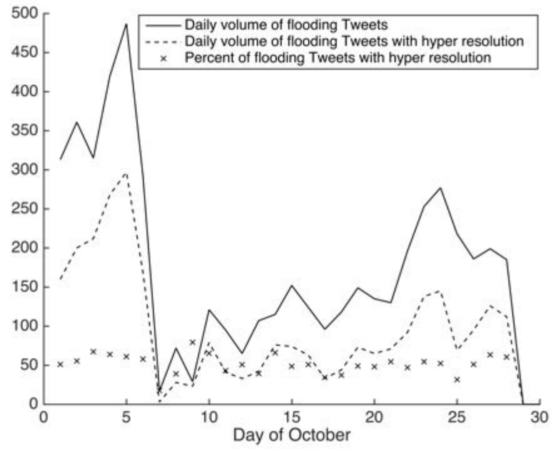


Figure 3 The daily volume of flooding related tweets and flooding tweets with hyper-resolution location information in the United States. The cross marks are the percentage of the hyper-resolution information among all the flooding Tweets.

Table 5 shows the number of identified flooded roads in the top 5 states. The top 3 numbers are consistent with the states that were impacted by Hurricane Joaquin. California (CA) and Texas (TX) are listed next to these states, because they are large and populate with high volume of tweets.

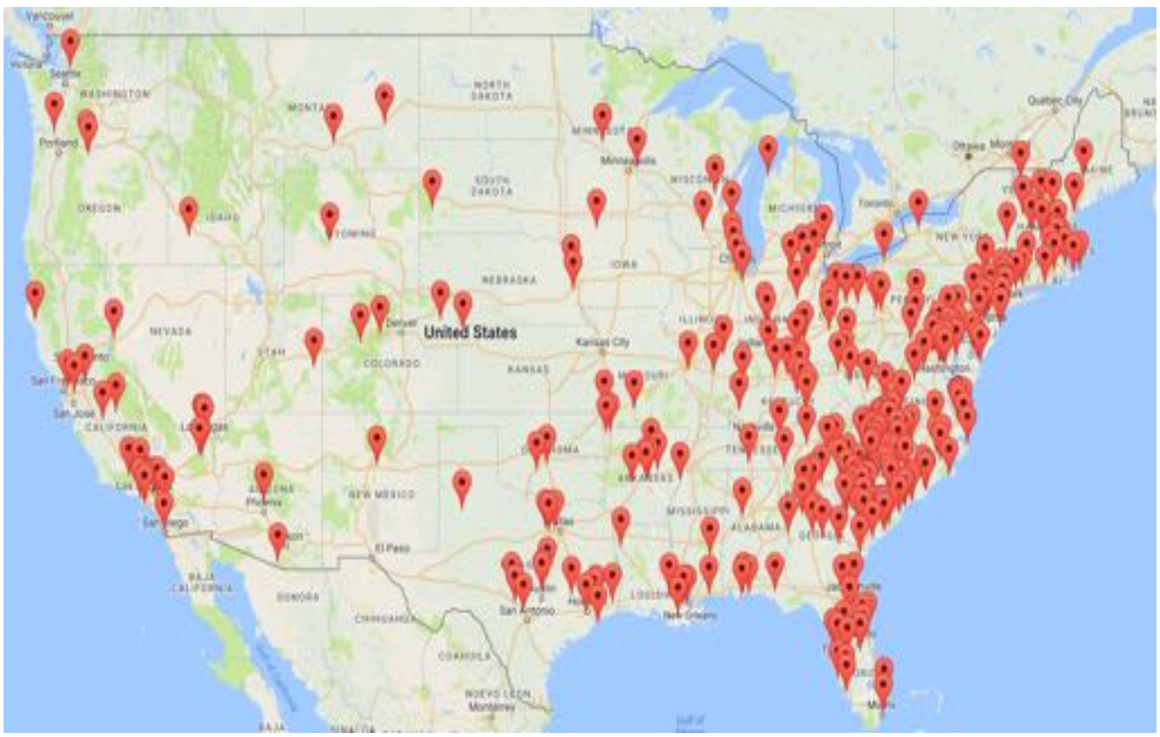
Table 5 Number of flooded roads by states: top 5 states.

State code	Number of roads
SC	57
NC	41
FL	21
CA	21
TX	17

The geo-spatial distribution of the hyper-resolution flooding tweets are shown in Figure 4. The identified tweets concentrated at the east coast capturing the extensive floods caused by the two hurricanes. The density of the tweets decreases toward Midwest but increases at the west coast and Gulf of Mexico.

To examine the reliability of the Twitter based flooding monitoring, we compute the correlation between the precipitation pattern and tweet volume. Using a grid of 50×50 cells, Figure 5 compares the geospatial patterns of precipitation observation, the precipitation departure, the percentage of precipitation departure, and the tweet volume. The precipitation data was obtained by averaging the data within the cell. The tweet volume was calculated by counting the tweets in each cell. A common feature among all the patterns is the high concentration around South Carolina, which captures the hurricane caused flooding. The tweet volume pattern has a high value in the stripe of Washington DC – New York – Boston, which is missing in all the other patterns. This mismatch indicates that urban areas that have dense population reported floods relatively more than remote places. This mismatch may bias the Twitter based flooding monitoring. Another mismatch is in Florida, where a high volume of flooding tweets is found there without heavy precipitation or precipitation departure. This finding may be attributed to the fact that the Florida floods were caused by tidal and storm surge. A high volume of flooding tweets is spotted in California, overlapping the precipitation-departure-percentage pattern. So we infer that the Californian floods during that period were caused by departure percentage of precipitation and California is more sensitive to the precipitation departure percentage than the absolute departure.

The Pearson correlation coefficient is used to examine the correlation between precipitation departure and tweet volume. We use different grids with the sizes of 100×100, 50×50, 25×25 and 12×12 to cover the contiguous United States. The Pearson correlation coefficient is calculated at different grid resolution in Table 6. The correlation coefficient is in the range of 0.17-0.41, so the tweet volume pattern has a weak linear correlation with the precipitation departure. Table 6 also shows that the correlation coefficient increases with coarse mesh, which suggests that Twitter based flooding monitoring is more reliable with high tweet volume.



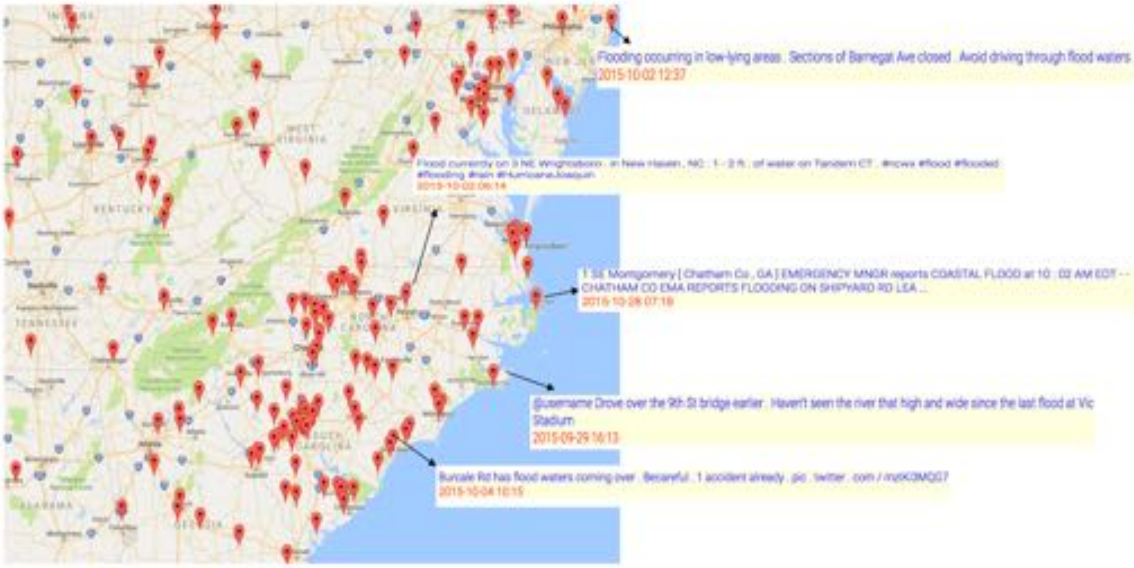
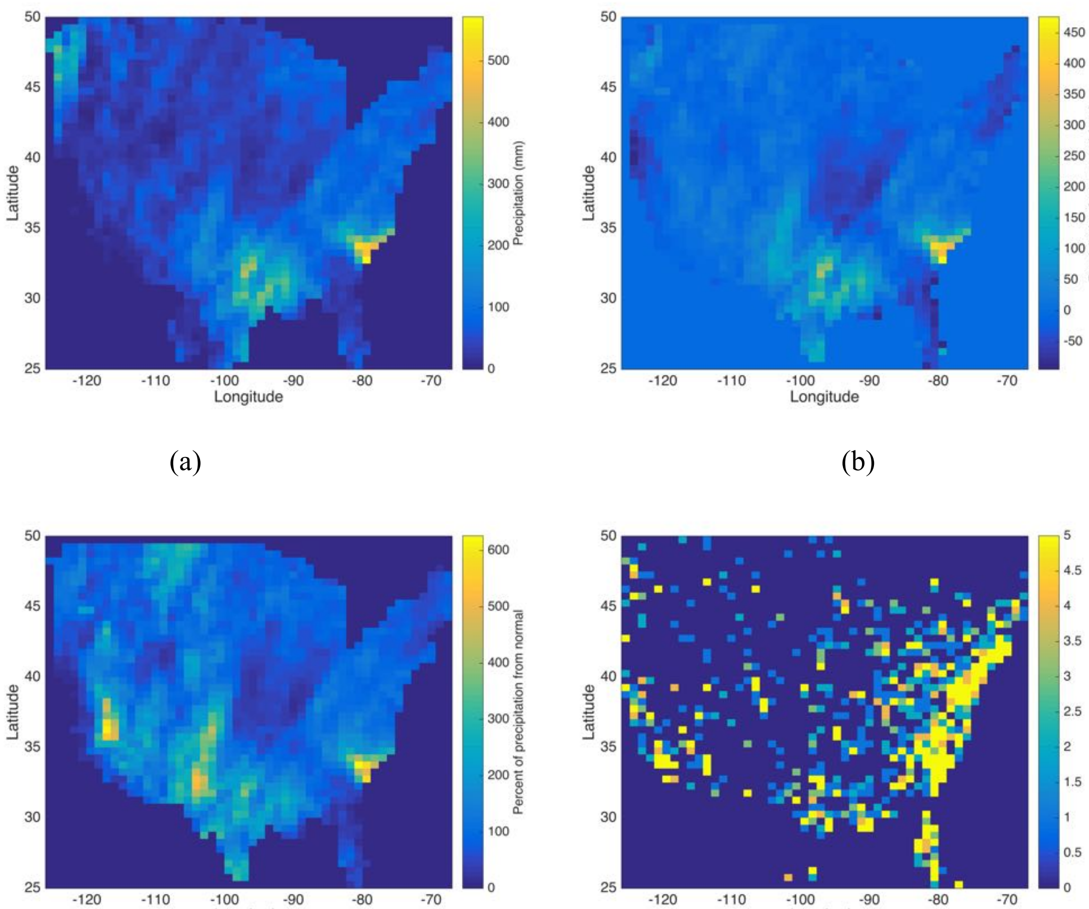


Figure 4 The spatial distribution of identified Tweets that have hyper-resolution geo-location information. Text content of sample posts is shown in the right panel.



(a) (b)

(c) (d)

Figure 5 Spatial patterns of precipitation and Tweets in October, 2015. (a) The observed precipitation. (b) The precipitation departure. (c) The percentage of precipitation departure. (d) The statistics of the Tweet volume.

Table 6 The Pearson correlation coefficients at different resolution of grids.

Grid size	Pearson correlation coefficient
100×100	0.1693
50×50	0.3156
25×25	0.3759
12×12	0.4095

The Ripley's K function is used to estimate the spatial pattern of the hyper-resolution flooding tweets. The Ripley's K function is a spatial analysis method to describe how point patterns cluster or disperse comparing to randomly distributed points over a given area of interest. The K function is defined (Dixon, 2002),

$$K(t) = A\sum_{i=1}^{n}\sum_{j=1}^{n} I_{t(i,j)} / n^2 \,, \tag{1}$$

where A is the area covered by the points, i and j are the indices of the points, n is the sample size, I_t is the impact function, which is one if the distance between points i and j is less than distance t and zero otherwise. A more commonly used quantity is the L function defined by

$$L = \sqrt{K(t)/\pi} - t \,. \tag{2}$$

The positive value of the L function means the data points are more clustered than a random process and the negative value means more dispersed. The L function is plotted against the distance t in Figure 6. The L maximum of around 190 km indicates that the data points are clustered the most at this length scale. It is consistent with the length scale of urban areas in the United States, e.g. the length scale of New York metropolitan area is around 185 km (Wikipedia, 2017, https://en.wikipedia.org/wiki/New_York_metropolitan_area). This indicates that the twitter pattern follows the population distribution in the US as expected. The L function drops to zero after the length scale of 100 km. So at this length scale the twitter posts distribute close to random pattern.

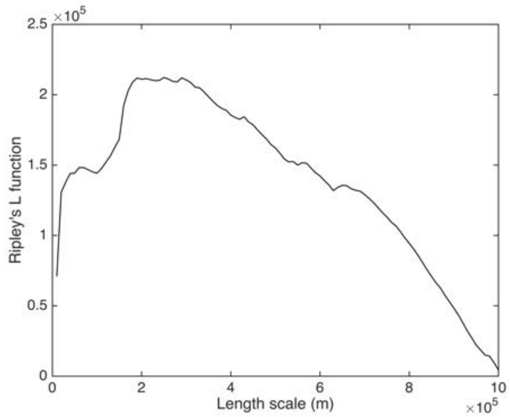


Figure 6 The Ripley's L function against the length scale *t*.

4.2 Case Study 2: Tidal flood at Charleston, SC

The most data collected through MyCoast is on September 29, 2015 in Charleston, SC, so we used this dataset as a demonstration of our flood monitoring technique. The processed crowdsourcing data as well as the road closure data are shown in Figure 7. Crowdsourcing data concentrated along the coastline of Charleston and most of the places were identified as "no flood" by the CV algorithm. The road closure data are distributed relatively inland.

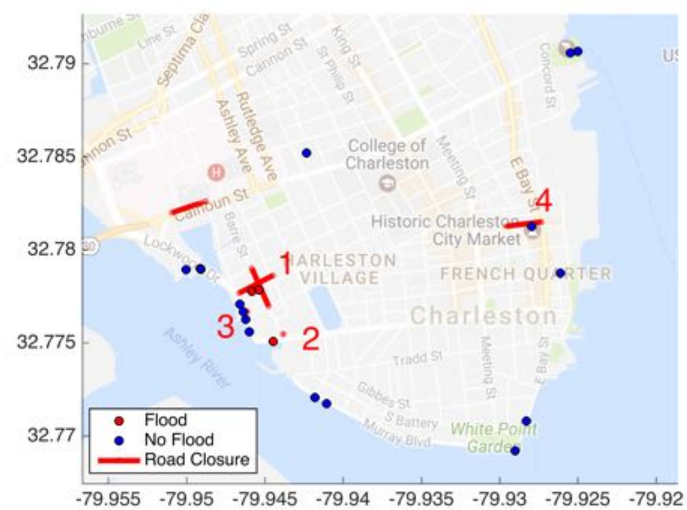


Figure 7 Comparison between crowdsourcing results and road closure data

The processed crowdsourcing data can be validated against the road closure data at four spots, because the two datasets roughly overlap in these locations, which are marked in Figure 7. The figure shows that crowdsourcing data correctly recognized the flood events at Spot 1 and 2, but made mistakes in Spot 3 and 4. The photos collected at Spot 3 are shown in Figure 8. The first and third photos show road floods, but the CV algorithm misclassified them. These mistakes could be explained by the strong reflection in the first photo and the semi-submerged plants in the third photo. These features resemble the natural water bodies and pose difficulties to process. The second photo shows no road flooding, but the flood water has reached a bench which should not be threatened by tidal floods. Again, the semi-submerged plants might confuse the CV. The last photo shows an overflowing drainage inlet. This photo indicates an overwhelmed drainage system and a nearby flood, but no road flooding was shown. So the CV code should not be blamed. The crowdsourcing photo collected at Spot 4 is shown in Figure 9. The floodwater has a regular edge, so the computer vision considered it as a normal water canal in the city. CV algorithm is proved useful to extract information, but the comparison shows there is a room to improve.

Figure 8 Photos near Road closure #3. From left to right, the photos are corresponding to the four points of Spot 3 in Figure 7 from South to North.

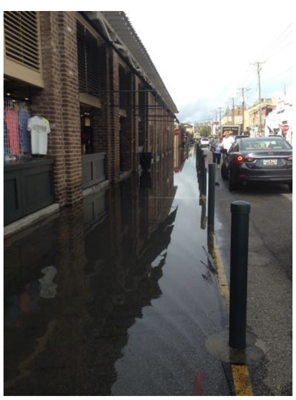


Figure 9 Photo near Spot 4.

5. Discussion

Comparing the two big-data approaches, we found they have different characteristics that potential users have to be cautious before direct use. Both methods can provide hyper-resolution monitoring with the resolution of street and parcel scale. Crowd-sourcing data is more advantageous because the GPS location can be up to the accuracy of meters, while the Twitter based data is in the scale of street names. In terms of geolocation coverage, Twitter based approach might be more favorable, because a much wider space and a worldwide monitoring could be feasible with little cost. However, we failed to locate any tweet at Charleston SC on September 29, 2015 to be compared with our road closure data. This mismatch indicates that

Twitter might be more suitable for large-scale monitoring. Crowdsourcing is restricted to the number of volunteers and distribution of app users, so its coverage can be specified in particular locations, but it might be still inappropriate to conduct large-scale monitoring. Regarding to the data volume and data collection speed, Twitter shows its advantage, so it may be more appropriate for real-time monitoring and higher volume of data benefits the accuracy of Twitter based monitoring.

An important issue that limits the big-data platform in practical use is data reliability. Crowdsourcing provides rich and customized information through its photos and user interface and the mobile phone app has high accuracy GPS sensors, so the collected data may have less uncertainty. In comparison, Twitter based approach has significant noise, which needs data cleaning before processing. However, the processing reliability might be opposite, for NLP is more accurate than CV at present. We expect higher reliability in the information extracting process from Twitter and other social media. Considering both of data collection and information extraction, in the case of a large quantity of data and automatic processing algorithm has to be involved, the general data reliability will be case by case, depending on the quality of data and training techniques of the practitioner. If a small amount of data needs to be processed, manual reading of crowdsourcing photos could be conducted, then crowdsourcing approach might be better. This might be the reason that some previous studies used photo data as the ground truth to validate numerical models.

On the subject of big data quality improvement, we outline some available measures and point to a series of ongoing research that has a hope to generate promising data quality control tools. For the crowdsourcing based approach, a photo shooting guidance should be provided so that a more complete scene can be captured and more reference objects can be included. A CV code should be retrained using the labeled crowdsourcing data to achieve high reliability. An emerging research trend is to determine water depth from the crowdsourcing photos by comparing to Google Street View. By detecting the difference between normal street view and the flooded streets, water depth can be inferred by calculating how much reference objects are blocked by floodwater. For Twitter based approach, in addition to the development of high accuracy NLP algorithm, a recent trend is to include “citizen science” in a feedback loop, so that a data elicitation can be sent to the original data contributor to request more information about the original report to pinpoint the location and follow the new development of the flood events. Water depth determined from the tweets are shown feasible in the present study, but its value has not been fully explored. In the future, water depth data should be compared against remote sensing data to cross-check the data quality. The photo contained in tweets can also be used to provide information about the flood. Fully exploiting such information can enrich the text-based monitoring effectiveness. A final note is that a good data fusion scheme, an approach to integrate of multiple data and knowledge and resolving data conflicts representing the same real-world object into a consistent, accurate, and useful representation, should be applied before using data to validate numerical models and support city planning and emergency response preparation (Liggins et al., 2017). This method has potential to greatly improve the data quality in general.

6. Conclusion

Urban flooding is difficult to monitor due to various complexities in data collection and processing. The present study shows that social media and crowdsourcing can be used to complement the datasets developed based on traditional remote sensing and witness reports. Applying these methods in two case studies, we found these methods are generally informative in flood monitoring. Twitter data is found weakly correlated to precipitation departure. We determined a length scale of tweet volume pattern, at which the data points are most clustered. The computer vision processed crowdsourcing data is compared against the road closure data. The results show that computer vision still has a room to improve, especially in coastal areas. These two methods are compared and a series of recommendation is given to improve the big data based flood monitoring in the future.

Acknowledgement

The first author thanks for the great support from Profs. Mark Stacey and Alexei Pozdnoukhov, and contributions from Wei Bai, Diyi Liu, Eason Ruan, Ruonan Ou, Renjie Wu, and Wenjun Zhong. The lead author also gratefully acknowledges the support of NSF, grant CBET-1541181. Huina Mao would like to thank the financial support for this research from the US government for Oak Ridge National Laboratory's Laboratory Directed Research and Development (project number LDRD 7677).

References

Eilander, D., Trambauer, P., Wagemaker, J. and van Loenen, A., 2016. Harvesting social media for generation of near real-time flood maps. *Procedia Engineering*, *154*, pp.176-183.

Finkel, J.R., Grenager, T. and Manning, C., 2005, June. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd annual meeting on association for computational linguistics* (pp. 363-370). Association for Computational Linguistics.

Fohringer, J., Dransch, D., Kreibich, H. and Schröter, K., 2015. Social media as an information source for rapid flood inundation mapping. *Natural Hazards and Earth System Sciences*, *15*(12), pp.2725-2738.

Gallien, T.W., Schubert, J.E. and Sanders, B.F., 2011. Predicting tidal flooding of urbanized embayments: A modeling framework and data requirements. *Coastal Engineering*, *58*(6), pp.567-577.

Guan, X. and Chen, C., 2014. Using social media data to understand and assess disasters. *Natural hazards*, *74*(2), pp.837-850.

Hess, L.L., Melack, J.M., Filoso, S. and Wang, Y., 1995. Delineation of inundated area and vegetation along the Amazon floodplain with the SIR-C synthetic aperture radar. *IEEE Transactions on Geoscience and Remote Sensing*, *33*(4), pp.896-904.

Hess, L.L., Melack, J.M., Novo, E.M., Barbosa, C.C. and Gastil, M., 2003. Dual-season mapping of wetland inundation and vegetation for the central Amazon basin. *Remote sensing of environment*, *87*(4), pp.404-428.

Jha, A.K., Bloch, R. and Lamond, J., 2012. *Cities and flooding: a guide to integrated urban flood risk management for the 21st century*. World Bank Publications.

Jongman, B., Wagemaker, J., Romero, B.R. and de Perez, E.C., 2015. Early flood detection for rapid humanitarian response: harnessing near real-time satellite and Twitter signals. *ISPRS International Journal of Geo-Information*, *4*(4), pp.2246-2266.

Lingad, J., Karimi, S. and Yin, J., 2013, May. Location extraction from disaster-related microblogs. In *Proceedings of the 22nd International Conference on World Wide Web* (pp. 1017-1020). ACM.

Liggins II, M., Hall, D. and Llinas, J. eds., 2017. *Handbook of multisensor data fusion: theory and practice*. CRC press.

Marcus, W.A. and Fonstad, M.A., 2008. Optical remote mapping of rivers at sub-meter resolutions and watershed extents. *Earth Surface Processes and Landforms*, *33*(1), pp.4-24.

Middleton, S.E., Middleton, L. and Modafferi, S., 2014. Real-time crisis mapping of natural disasters using social media. *IEEE Intelligent Systems*, *29*(2), pp.9-17.

Muralidharan, S., Rasmussen, L., Patterson, D. and Shin, J.H., 2011. Hope for Haiti: An analysis of Facebook and Twitter usage during the earthquake relief efforts. *Public Relations Review*, *37*(2), pp.175-177.

Nguyen, D.T., Alam, F., Ofli, F. and Imran, M., 2017. Automatic Image Filtering on Social Networks Using Deep Learning and Perceptual Hashing During Crises. *arXiv preprint arXiv:1704.02602*.

Rees, W. G. (2001). Physical principles of remote sensing. Cambridge, UK: Cambridge University Press.

Smith, R.A., Bates, P.D. and Hayes, C., 2012. Evaluation of a coastal flood inundation model using hard and soft data. *Environmental Modelling & Software*, *30*, pp.35-46.

Sun, D., Li, S., Zheng, W., Croitoru, A., Stefanidis, A. and Goldberg, M., 2016. Mapping floods due to Hurricane Sandy using NPP VIIRS and ATMS data and geotagged Flickr imagery. *International Journal of Digital Earth*, *9*(5), pp.427-441.

Wilson, J.N., Gader, P., Lee, W.H., Frigui, H. and Ho, K.C., 2007. A large-scale systematic evaluation of algorithms using ground-penetrating radar for landmine detection and discrimination. *IEEE Transactions on Geoscience and Remote Sensing*, *45*(8), pp.2560-2572.

Zeiler, M.D. and Fergus, R., 2014, September. Visualizing and understanding convolutional networks. In *European conference on computer vision* (pp. 818-833). Springer International Publishing.